Analysing Student VLE Behaviour Intensity and Performance

Jakub Kuzilek[1(✉)], Jonas Vaclavek[1], Zdenek Zdrahal[1,2], and Viktor Fuglik[1,3]

[1] CIIRC, CTU in Prague, Jugoslavskych partyzanu 1580/3, 160 00 Prague, Czech Republic
jakub.kuzilek@cvut.cz

[2] KMi, Open University, Walton Hall, Milton Keynes MK7 6AA, UK

[3] Faculty of Education, Charles University, Magdaleny Rettigove 4, 116 39 Prague 1, Czech Republic

Abstract. Almost all higher educational institutions use Virtual Learning Environments (VLE) for the delivery of educational content to the students. Those systems collect information about student behaviour, and university can take advantage of analysing such data to model and predict student outcomes. Our work aims at discovering whether there exists a direct connection between the intensity of VLE behaviour represented as recorded student activities and their study outcomes and analyse how intense this connection is. For that purpose, we employed the clustering method to divide students into so-called VLE intensity groups and compared formed groups (clusters) with the student outcomes in the course. Our analysis has been performed using Open University Learning Analytics dataset (OULAD).

Keywords: Clustering · Virtual Learning Environment · Student performance · Predictive modeling

1 Introduction

At present, many higher education institutions already introduced ICT based online education systems in their portfolio. These Virtual Learning Environment systems such as Moodle platform [1] deliver educational content directly to students anytime and anywhere. This trend is further boosted by the introduction of Massive Open Online Courses (MOOCs) platforms.

Online educational platforms collect data about their users. VLEs together with other data sources commonly used by higher education institutions make it possible to analyse student data. Various methods of using data for education improvement have been investigated in more than 200 studies in recent years [2].

M. Scheffel et al. (Eds.): EC-TEL 2019, LNCS 11722, pp. 587–590, 2019.
https://doi.org/10.1007/978-3-030-29736-7_45

2 Research Question

The research aims at answering the question: *Is it possible to uncover natural grouping of students based on their VLE activities without prior knowledge of their results?* To answer this question, we employed expectation maximisation clustering[1] on VLE behaviour data available in Open University Learning Analytics dataset (OULAD) [3] and compare the created activity groups with the student assessments results. The reported results are part of the larger outgoing project at CTU in Prague.

3 Data

The OULAD [3] contains data of about 32,593 students studying 22 OU courses in years 2013 and 2014. The OU is the largest British distance learning university with approximately 170,000 students. The typical OU course has one or more assignments, final exam and has a length of about 35 weeks. OU uses the Moodle platform to provide learning materials to students. In addition, the system provides the framework for submitting assignments and their evaluation. For more details see the original paper [3]. The dataset includes data about both students and courses. We analyse data of the FFF course and the 2014 J presentation, which is one of the STEM subjects offered by the university. More than 1/3 of the students have withdrawn during the presentation.

4 Methods

To compare the student's VLE behaviour with their performance in assessments, it is necessary to transform VLE logs and to adjust "performance classes" based on student scores in assessments.

4.1 VLE Behaviour Intensity and Assessment Performance

At first, all VLE log entries from the time prior to the start of the course have been filtered out. Those represent outliers, and their added value in this task is minimal. Next, we transformed daily VLE logs into the weekly aggregates. Keeping the information about how many times the student clicked into the VLE system every week makes data less sparse and more robust against spikes of activities. The summary number of weekly clicks is considered as a measure of VLE behaviour intensity.

For the analysis, we need to adjust the student assessment scores and create performance groups. For that purpose, the scores ranging from 0 to 100 are divided into six possible performance classes: **Not submit** (the student did not submit the assessment); **Submitted and failed** (student failed with score less than 40 points); **Lowest**

[1] L. Scrucca, M. Fop, T. B. Murphy and A. E. Raftery, "mclust 5: clustering, classification and density estimation using Gaussian finite mixture models," *The R Journal*, 2016.

passing score (student scored 41–55 points); **Low passing score** (student scored 56–70 points); **Medium passing score** (student scored 71–85 points); **High passing score** (student scored more than 86 points). We assigned numbers 1 (Not submit) – 6 (High passing score) to student performance classes.

4.2 Clustering Student VLE Behaviour Data

Student VLE behaviour intensity forms the dataset for unsupervised learning. For that purpose, we employed Gaussian finite mixture models fitted via the EM algorithm. The resulting model then produced a set of labels which can be further compared with the assessment performance classes to explore whether the student behaviour intensity relates to the assessment performance classes. In our research, we set the number of clusters to 6 to keep the comparison simple.

4.3 Comparing Clusters and Assessment Performance

We are interested in "overlaps" between the clusters created by Gaussian finite mixture models and assessment performance classes. Thus the type of contingency can be created, which element x_{ij} represents the proportion of students from cluster i belonging to assessment performance class j.

Table 1. Comparison of VLE behaviour intensity based clusters and assessment performance

	Class	1	2	3	4	5	6	1	2	3	4	5	6
		Assessment 1						Assessment 2					
Cluster	1	1	0	3	22	43	31	5	4	7	20	37	28
	2	0	0	1	16	43	39	0	1	3	14	33	49
	3	34	3	6	23	25	9	86	4	4	5	2	0
	4	6	0	5	30	40	20	35	7	8	23	17	10
	5	0	1	2	11	40	47	0	1	4	11	25	60
	6	90	0	1	4	5	1	91	0	1	5	2	0
		Assessment 3						Assessment 4					
Cluster	1	11	8	10	18	37	16	22	7	13	19	28	11
	2	1	3	5	18	45	28	2	4	8	16	41	29
	3	100	0	0	0	0	0	100	0	0	0	0	0
	4	85	6	1	5	2	1	95	2	2	1	0	0
	5	2	2	5	10	43	39	4	2	6	12	32	44
	6	96	0	0	3	0	1	99	1	0	1	0	0
		Assessment 5											
Cluster	1	43	9	6	14	16	11						
	2	9	4	5	14	32	37						
	3	100	0	0	0	0	0						
	4	100	0	0	0	0	0						
	5	9	3	2	9	24	53						
	6	99	0	1	0	0	0						

5 Results and Discussion

Table 1 contains the results of a comparison of student VLE behavioural clusters with the assessment performance classes for all assessments in the course *FFF*.

One can observe that clusters can be divided into those with high assessment performance and those with low assessment performance. From the very first assessment, one can find that the majority of students in cluster 6 are not going to submit any assessment and this cluster can be viewed as the lowest performing cluster. Cluster 3 is the cluster with second lowest performance, and these students tend to give up after submitting their first assessment. Cluster 4 is formed by students who tend to give up after the second assessment. On the other hand, clusters 2 and 5 consists of students who have the highest performance in the assessments. Cluster 5 is containing the best students, which can be viewed especially in the second and fifth assessment. The "average" students fall into cluster 1. These students are uniformly distributed at the beginning, and when time progresses, they perform slightly worse.

6 Conclusion

In this paper, we analysed VLE data of one OU course with the expectation maximization algorithm to answer the question whether the student activities form "performance" groups. The formed clusters of students based on behavioural intensity were compared with student's performance in their assessments. The comparison shows that even if data does not contain the information about the outcomes, one can still efficiently analyse and detect groups of students at risk of failure. For example, there exists clear group of students who failing from the very beginning of the course. In overall, one can observe that the results of students same as students VLE activity drops. We plan to further extend this research by a deeper analysis of formed clusters to better understand the phenomena lying behind the formed behavioural groups. For example, the comparison of average VLE intensity and assessment scores will give us insight to the relationship between activities in VLE and assessments.

Acknowledgement. This work was supported by junior research project no. GJ18-04150Y and student research grant no. SGS19/209/OHK3/3T/37.

References

1. Moodle HQ: Moodle statistics. Moodle HQ (2018). https://moodle.org/stats/. Accessed 25 Apr 2018
2. Papamitsiou, Z., Economides, A.A.: Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. Educ. Technol. Soc. **17**, 49–64 (2014)
3. Kuzilek, J., Hlosta, M., Zdrahal, Z.: Open university learning analytics dataset. Sci. Data **4** (2017)